## INTERVIEW SUMMARY

Applicants would like to thank Examiner Saint Cyr and his supervisor for the courtesies he extended them and their attorney in the personal interview conducted on October 23, 2006. In the interview Applicants' attorney discussed the patentability of, *inter alia*, dependent claims 9, 10, 17 and 28 indicating that the citations to the Gillis reference (US 6,523,026) in the Office Action did not teach their respective features. The Examiner agreed to review the Gillis reference as to these claims. Further, the Examiner agreed that as to dependent claims 15, 25 and 32, these claims distinguished over Gillis. These claims have been rewritten as new independent claims.

## REMARKS

In view of the foregoing amendments and following remarks, reconsideration and allowance of this application is respectfully requested. Claims 1-32 stand rejected. Claims 1, 6, 7, 10, 11, 14, 16, 23, 24, 26 and 27 have been amended. Claims 8-9, 15, 25 and 32 have been canceled without prejudice, and new claims 33-35 are presented, which are dependent claims 15, 25 and 32 rewritten as independent claims. Claims 1-7, 10-14, 16-24, 26-31 and 33-35 are pending in this application. No new matter has been introduced.

Independent claims 1, 14, 23, 24, 26 and 27 and dependent claims 2-13, 16-22, and 28-31 stand rejected under 35 U.S.C. § 102(a) as being anticipated by Gillis (U.S. Patent No. 6,523,026). Applicants respectfully traverse these claim rejections for the reasons set forth hereinafter.

As set forth in detail in the present application, Applicants' claimed invention is directed to a system and method for the comparative analysis of textual documents including accessing two or more documents, performing a linguistic analysis on each document, outputting a quantified representation of a semantic content of each document, and comparing the

quantified representations using a defined metric. Such a metric can measure relative semantic closeness or distance of two documents. The semantic content can be presented as a semantic vector, the format of which can be flexible enough to allow any metric used to operate on it to be adapted and optimized to the type and/or domain of documents being analyzed and the goals of the comparison.

The semantic vector as claimed in claim 1 comprises multiple components, wherein each component of said semantic vector has at least a word or phrase appearing in the document or a synonym of said word or phrase, a weighting factor associated with said word or phrase or synonym, and a frequency value.

The defined metric of independent claims 14, 23, 24, 26 and 27 measures the semantic distance between two semantic vectors as a function of the relative frequencies of one of common terms and common {main term-subordinate term pairs} between the two documents.

The Gillis patent cited by the Examiner describes embodiments for a process of identifying terms or sets of terms in target domains having functional relationships analogous to terms selected from a source domain whereby query relevant, but semantically distant, analogies may be retrieved, corresponding to any user defined query. The method disclosed by Gillis comprises the automated generation of an abstract representation of terms from a first user selected knowledge domain, the representations encoding the co-occurrence patterns of terms characteristic of the source domain, and application of the representations to the efficient discovery of analogous objects in one or more semantically distant target domains. The abstract representations are most preferably vectors in a high dimensionality space. A small subset of terms is chosen from the source domain, the terms in such subset being substantially absent from the target domains and having substantially no known equivalents in the target domains. These

source domain specific terms are those for which analogous objects are sought in the target domains. These analogous objects are terms or groups of terms from the target domains, which are in some way related to the chosen source domain terms.

Gillis does not teach or suggest outputting a quantified representation of the semantic content of each document and comparing the quantified representations using a defined algorithm or metric. Rather, Gillis attempts to compare semantic similarity of terms by computing vector overlap. Specifically, Gillis does not teach comparing the quantified representations using a defined algorithm or metric, wherein the defined algorithm or metric is one of:

$$[\mathrm{Sqrt}(f1^2 + f2^2 + f3^2 + f4^2 + \ + f(N\text{-}1)^2\, fN^2)/n] \ * 100,$$

wherein f is a difference in frequency of a common term between two documents and n is the number of terms those documents have in common; or

Sqrt(sum((w-Delta)^2 * w-Avg)) / (Log(n)^3 * 1000),

wherein w-Delta is the difference in weight between two common terms, w-Avg is the average weight between two common terms, and n is the number of common terms, between two documents.

Thus, as discussed in the personal interview, new claims 33-35 are patentably distinguishable over Gillis.

Additionally, as discussed in the personal interview, Gillis does not teach comparing the quantified representations using a defined algorithm or metric, wherein the defined algorithm or metric measures the semantic distance between two semantic vectors as a function of the relative frequencies of one of common terms and common {main term-subordinate term pairs} between the two documents. Gillis is wholly silent as to the

phenomenon of {main term-subordinate term pairs}in general, and in particular as to its use as an indicator of semantic similarity over and above simply analyzing common terms, as is discussed in detail in Specification at Appendix V, ¶¶ 548, 553, 558 and 566, for example.

Thus, as submitted in the personal interview, independent claims 14, 23, 24, 26 and 27 are patentably distinguishable over Gillis.

Finally, claim 1 provides that a semantic vector has multiple components, wherein each component of said semantic vector has at least: (i) a word or phrase appearing in the document or a synonym of said word or phrase, (ii) a weighting factor associated with said word or phrase or synonym, and (iii) a frequency value.

Gillis does not teach a semantic vector whose components each have at least these three values. In particular, the only reference in Gillis to frequency of a term is at 41:39-56. There the concept "inverse frequency weighting" is introduced. Under such a method, "vectors for ***terms*** that occurred less frequently in the training corpus are weighted more heavily in the calculation of summary vectors of search domain records and queries." *Id.* at 41:43-46. The claimed invention does not utilize inverse frequency weighting, and the claimed invention does not generate vectors for **terms**. The claimed invention generates semantic vectors for **documents**. The semantic vectors of claim 1 have components, and each component has at least three values: a word or phrase, a weighting factor for said word or phrase, and a frequency value for said weighted word or phrase (it being recalled that different words can be weighted differently depending upon their context of use in a document; thus there can be multiple components of a semantic vector for the same word). Frequency values appear as a third value in each component of each semantic vector of claim 1. They are orthogonal to any weighting of said word or phrase, unlike the scheme of Gillis. Because Gillis uses frequency solely to

(inversely) weight a term vector, Gillis cannot provide frequency as a separate and independent value in a semantic vector. Gillis simply does not teach or suggest using frequency as an independent value of a component of a semantic vector.

Thus, independent claim 1 is patentable over Gillis.

The remaining dependent claims are also allowable for similar reasons.

No additional fee is believed to be due with the filing of this Amendment. However, if any fee is due, the Director is hereby authorized to charge any additional deficiencies or credit any overpayments to Deposit Account No. 50-0540.

Dated: November 19, 2007

Respectfully submitted,

By: _____

Aaron S. Haleva, Reg. No. 44,733
KRAMER LEVIN NAFTALIS & FRANKEL LLP
1177 Avenue of the Americas
New York, New York 10036
(212) 715-7773 (phone)
(212) 715-9397 (fax)